



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

# Genetic drift and within-host metapopulation dynamics of HIV-1 infection

### Citation for published version:

Frost, SDW, Dumaurier, MJ, Wain-Hobson, S & Leigh Brown, AJ 2001, 'Genetic drift and within-host metapopulation dynamics of HIV-1 infection', *Proceedings of the National Academy of Sciences (PNAS)*, vol. 98, no. 12, pp. 6975-6980. <https://doi.org/10.1073/pnas.131056998>

### Digital Object Identifier (DOI):

[10.1073/pnas.131056998](https://doi.org/10.1073/pnas.131056998)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

Proceedings of the National Academy of Sciences (PNAS)

### Publisher Rights Statement:

Free in PMC.

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Genetic drift and within-host metapopulation dynamics of HIV-1 infection

Simon D. W. Frost<sup>\*†</sup>, Marie-Jeanne Dumaourier<sup>‡</sup>, Simon Wain-Hobson<sup>‡</sup>, and Andrew J. Leigh Brown<sup>\*§</sup>

<sup>\*</sup>Department of Pathology, University of California, San Diego, CA 92103; <sup>†</sup>Unité de Rétrovirologie Moléculaire, Institut Pasteur, 75724 Paris Cedex 15, France; and <sup>§</sup>Centre for HIV Research, University of Edinburgh, Edinburgh EH9 3JN, Scotland

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved April 11, 2001 (received for review February 5, 2001)

**Most HIV replication occurs in solid lymphoid tissue, which has prominent architecture at the histological level, which separates groups of productively infected CD4<sup>+</sup> cells. Nevertheless, current population models of HIV assume panmixis within lymphoid tissue. We present a simple “metapopulation” model of HIV replication, where the population of infected cells is comprised of a large number of small populations, each of which is established by a few founder viruses and undergoes turnover. To test this model, we analyzed viral genetic variation of infected cell subpopulations within the spleen and demonstrated the action of founder effects as well as significant variation in the extent of genetic differentiation between subpopulations among patients. The combination of founder effects and subpopulation turnover can result in an effective population size much lower than the actual population size and may contribute to the importance of genetic drift in HIV evolution despite a large number of infected cells.**

HIV has an enormous evolutionary potential because of the combination of a high mutation rate ( $\approx 3 \times 10^{-5}$  substitutions/site/generation; ref. 1), a short generation time ( $\approx 2.6$  days; refs. 2 and 3), and a large number of infected cells within the host ( $10^7$ – $10^8$  infected cells; ref. 4). Many analyses of the pattern of nucleotide substitution in the viral *env* gene have detected evolution in response to positive selection (5–16), although relatively few studies have linked these patterns of viral genetic variation to selection by defined immune responses or for changes in viral phenotype associated with coreceptor usage. In the presence of antiretroviral agents that target viral protease and reverse transcriptase, positive selection has been described in numerous studies, which demonstrate convergent evolution in different individuals of mutations that confer drug resistance (17–20). Although specific mutations have been identified with escape from immune responses and escape from antiretrovirals, viral genetic variation reflects more than just these focused positive selection pressures and may reflect more subtle processes of within-host competition between viruses for higher replicative potential.

Despite an enormous potential to respond to selection, there can be substantial differences between infected individuals in the evolution of drug resistance. For example, the frequency of single drug-resistant mutants in viral protease and reverse transcriptase before therapy can vary dramatically between hosts (21–23), apparently by random. Indirect observations of the apparently stochastic way in which multiply resistant mutants emerge during therapy are also consistent with a role of genetic drift (18, 24, 25). In addition, three studies of the evolution of *env* gene sequences have found substantially lower genetic variation than expected for a population of the order of  $10^7$  (24, 26, 27). These studies estimated the “effective” population size to be of the order of  $10^3$ , reflecting the high relatedness between *env* sequences sampled at a given time. On the basis of the level of variation in the frequency of M184V mutation in reverse transcriptase, Frost *et al.* (23) estimated an effective population size of  $10^5$ – $10^6$ , similar to with estimates of  $10^4$ – $10^5$  estimated from patterns of variation in protease (28). Hence, stochastic models of drug resistance evolution that assume an effective population

size close to the actual population size (29) may dramatically underestimate the role of genetic drift in generating variation between individuals in the emergence of drug-resistant strains.

Despite empirical evidence for an effective population size low enough to generate considerable variation in the frequency of rare drug-resistant mutants, the means by which a large actual population size is translated into an effective population size at least 10- to 100-fold smaller are not understood and reveal a significant gap in our understanding of viral dynamics. Although many models of HIV-1 evolution assume that the population of infected cells within the host is well-mixed (28, 30), there is direct evidence that this is an unrealistic assumption. On a macroscopic spatial scale, it has been recognized for some time that HIV genetic variability is spatially heterogeneous (31–36). Studies of proviral clones isolated from individual splenic white pulps have revealed that such differentiation can also occur on a microscopic scale (37). Despite many potential target cells for HIV infection in splenic tissue, the average number of infected cells is low, with a high frequency of putative recombinants, suggesting high levels of mixing at a local scale but poor mixing of virus at larger spatial scales (38).

To take into account the evidence for population structuring, we have applied a simple “metapopulation” model (39) of HIV replication, in which small subpopulations of infected cells are established by a small number of founding viruses and turn over at a high rate. To test this model, we analyzed the pattern of genetic variation in the V1/V2 region of the *env* gene within and between infected white pulps in the spleens of six HIV-1-infected individuals. Founder effects are shown to play an important role in determining the effective rate of recombination, and that subpopulation turnover can result in an effective population size that is much lower than the number of infected cells in the body, thus contributing to the action of genetic drift in the evolution of HIV populations.

## Methods

**Sequence Data.** White pulps were microdissected from spleens and  $\approx 20$  proviral clones of the V1/V2 region of *env*  $\approx 300$  bp long were sequenced from two to five pulps per spleen, as described elsewhere (ref. 37; M.-J.D. and S.W.-H., unpublished work). Sequences were aligned with CLUSTALW (40) and checked by eye. Although in peripheral blood, proviral DNA from peripheral blood mononuclear cells (PBMCs) can differ from viral RNA in plasma, presumably because of the “archiving” of genetic variation in PBMCs over time (3, 5), this is unlikely to be an issue here, as (M.-J.D. and S.W.-H., unpublished work) have shown that viral RNA and proviral DNA obtained from the same pulp are virtually indistinguishable.

This paper was submitted directly (Track II) to the PNAS office.

<sup>†</sup>To whom reprint requests should be addressed at: Department of Pathology, School of Medicine, University of California at San Diego, UCSD Treatment Center, 150 W. Washington Street, Suite 100, San Diego, CA 92103. E-mail: sdfrost@ucsd.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

**Population Genetic Analysis.** Genetic differentiation between subpopulations was quantified by using estimates of  $F_{ST}$ , the fraction of the total genetic variation found between subpopulations (41) obtained from an analysis of molecular variance (AMOVA) in ARLEQUIN ver. 1.1 (42) assuming a Jukes–Cantor (43) model to correct for multiple hits. The significance of departures from the null hypothesis of a random distribution of genetic variation was determined by using 10,000 randomizations of sequences between populations.

Estimates of the number of pairwise differences and the number of mutations for each sample were obtained by using DNASP ver. 3.0 (44). Tajima's  $D$  statistic was calculated by using a  $\beta$  distribution with mean 0 and variance 1 to estimate the  $P$  value against the values of  $D$  expected under a coalescent model of a population of constant size (45). Values of  $D \leq 1.84$  give a  $P$  value of approximately less than 0.05 when the sample size is  $\approx 20$  (46).

$F_{ST}$  and Tajima's  $D$  were calculated, correcting for heterogeneity in the substitution rate across sites (47). Rate variation was modeled as a  $\gamma$  distribution with shape parameter 0.3, similar to the values of the shape parameter estimated from the V3 region of *env* (48).

We tested for recombination within each individual by testing for long regions of sequence that are sufficiently long and/or similar to suggest recombination either between sequences in the sample or their ancestors ("inner" fragments) or between a sequence in the sample and another sequence that was unknown ("outer sequence" fragments), either because the other sequence was not present in the sample or because further mutation had obscured the relationship between the two sequences. Global  $P$  values, corrected for multiple comparisons, were obtained by simulation using the program GENECONV ver. 1.81 (49). To compare the extent of recombination between sequences within each pulp, the minimum number of recombination events was estimated by analyzing pairs of variable sites by using the method of Hudson and Kaplan (50). Sites at which three or four nucleotides were segregating were not included.

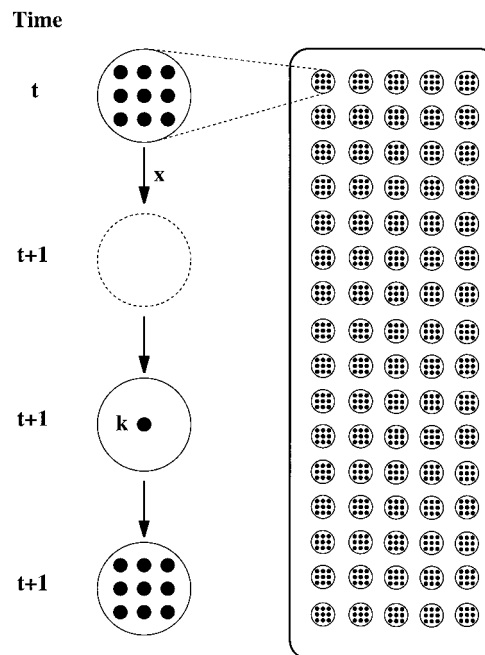
## Results

To develop a biologically plausible model of HIV replication within solid tissue, a simple model of molecular evolution in a metapopulation was considered as originally studied by M. Slatkin (51) and recently revisited in the context of nucleotide diversity by Pannell and Charlesworth (52). In this model, the total number of infected cells in an individual,  $N_T$ , remains constant and consists of  $n$  subpopulations (where  $n \gg 1$ ), each made up of  $N_S$  infected cells,  $N_T = nN_S$ . A proportion of subpopulations, denoted  $x$ , goes extinct every generation and is replaced by new subpopulations, each of which are founded by  $k$  infected cells drawn randomly from the whole population. It is assumed that the probability of founders being drawn from the same subpopulation is zero. This assumption is intended to reflect long-range migration from a well-mixed pool of latently infected cells to pockets of uninfected cells, as hypothesized by Cheynier *et al.* (37). For simplicity in dealing with subpopulation age structure, it was assumed that the  $k$  founders immediately grow into a subpopulation of  $N_S$  infected cells and that generations are discrete. A schematic of the model is given in Fig. 1.

At equilibrium,  $F_{ST}$ , defined as the amount of genetic variation between subpopulations expressed as a fraction of the total variation (41), is given by the expression.

$$F_{ST} = \frac{k + N_S x}{k(1 + N_S x)} \approx \frac{1}{k} \quad [1]$$

The approximation above is for moderate amounts of subpopulation turnover,  $x \gg 1/N_S$ . For example, for  $N_S = 100$ , the approximation is valid if the average lifetime of a subpopulation



**Fig. 1.** Schematic of the metapopulation model. The total pool of HIV-infected cells,  $N_T$ , is assumed to consist of a large number,  $n$  (for illustrative purposes,  $n = 16 \times 5$ ), of subpopulations, each consisting of  $N_S$  cells (here,  $N_S = 9$ ), as shown on the *Right*. A proportion  $x$  of subpopulations go extinct every generation, and new subpopulations are established in their place by  $k$  founders (here,  $k = 1$ ), which for mathematical simplicity are assumed to increase to  $N_S$  infected cells instantly, as shown on the *Left*.

is less than 100 generations ( $\approx 200$  days). Given the low average frequency of infected cells within splenic tissues, with  $<1\%$  of  $CD4^+$  T cells harboring proviral DNA (37), subpopulation turnover is likely to be much higher than this.

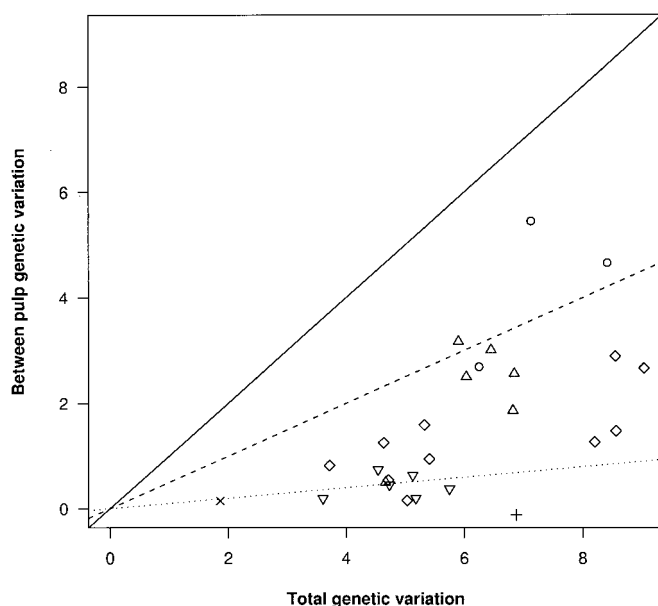
To test whether the metapopulation model provides a better fit to the pattern of viral genetic variation in solid tissues than a panmictic model, sequence data of the V1/V2 hypervariable region of the *envelope* gene obtained from two or more white pulps each from the spleens of six HIV-1-infected individuals were analyzed (ref. 37; M.-J.D. and S.W.-H., unpublished work).  $F_{ST}$  was calculated for each spleen, both using all of the pulps, to get an overall measure of between host variation in genetic differentiation (Table 1), as well as for each pair of pulps, to indicate potential within-host variation in genetic differentiation (Fig. 2). Estimates of  $F_{ST}$  allowing for heterogeneity in the substitution rate along the sequence were slightly higher but were very similar (results not shown).

With the exception of patient M, where the estimate of between-pulp variation was negative, reflecting sampling error

**Table 1. Results of analysis of molecular variance (AMOVA) for each patient**

Patient	Number of pulps analyzed	Between-pulp variation $V_a$	Total variation $V_b$	$F_{ST}$ ( $=V_a/V_b$ )
B	3	1.453	2.446	0.594
L	4	0.836	2.266	0.369
M	2	-0.041	2.380	-0.017 <sup>ns</sup>
N	2	0.055	0.706	0.078
P	5	0.463	2.154	0.215
S	4	0.146	1.629	0.090

ns, not significant at the 5% level.

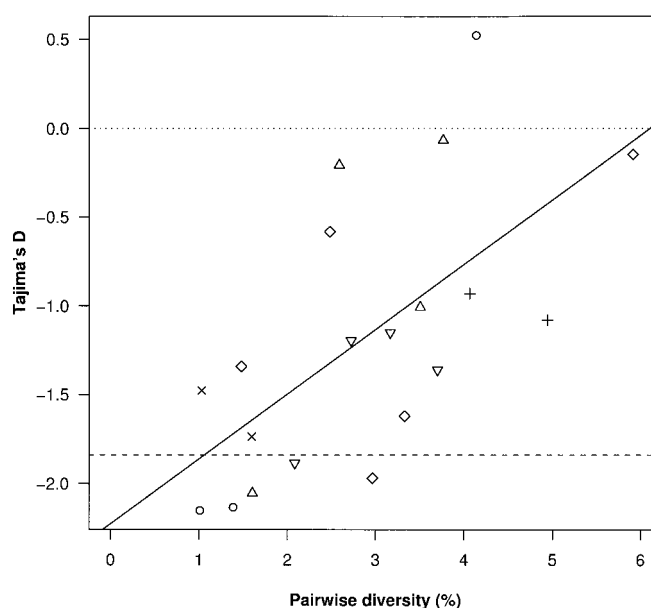


**Fig. 2.** Plots of genetic variation (as measured by the variance components in an analysis of molecular variance) between pairs of infected pulps against total genetic variation. Lines indicate contours of  $F_{ST}$ : 0.1 (dotted line), 0.5 (dashed line), and 1 (solid line). Data from different individuals are represented by different symbols (B =  $\circ$ , L =  $\Delta$ , M = +, N =  $\times$ , P =  $\diamond$ , S =  $\cdots$ ). With the exception of the pair of pulps from patient M, all values of  $F_{ST}$  are significantly greater than zero ( $P < 0.05$ ), indicating deviation from panmixis, although there is great variation in  $F_{ST}$  between pulps.

coupled with homogeneity of viral genetic variation across pulps,  $F_{ST}$  was significantly higher than expected, based on a random distribution of variation ranging from 0.08 for patient N to 0.59 for patient B (Table 1).  $F_{ST}$  for pairs of infected pulps was also higher than expected from a random distribution of variation (with the exception of a single comparison in patients P and S) and also varied dramatically between individuals, ranging from very low ( $<0.1$ ) to very high (0.76) levels of divergence between pulps (Fig. 2). Levels of genetic differentiation were rather similar between pulps within a patient, with the exception of patient P, which shows two clusters of points because of a single pulp with a very high ( $\approx 6\%$ ) diversity (Fig. 2). Hence the homogenous model was rejected. In addition, the metapopulation model can easily account for the high observed variation in  $F_{ST}$  between individuals, by subtle differences in the number of founders,  $k$ , as with moderate amounts of subpopulation turnover, such that  $N_{SX} \gg 1$ ,  $F_{ST}$  is approximately equal to  $1/k$ .

The metapopulation model predicts the occurrence of founder effects, random changes in frequency of genetic variants because of sampling when the number of founders are low. These effects can be detected by using Tajima's  $D$  statistic, which compares pairwise diversity with the number of variable sites in a sample of sequences (45). Pairwise diversity is an estimator of the evolutionary distance to the most recent common ancestor (MRCA), whereas the number of variable sites is an estimator of the overall length of the phylogeny. As founder effects result in phylogenetic trees with short distances to the MRCA (hence low pairwise distances) and with a star-like shape (hence many variable sites), founder effects can give rise to values of  $D$  less than 0.

Fig. 3 shows that there is a positive correlation between within-pulp diversity and Tajima's  $D$  (Spearman's rank correlation = 0.70,  $P < 0.001$ ), with a significant bottleneck effect when within-pulp diversity is low, as expected if low within-pulp diversity arises because of a low number of founders. Estimates

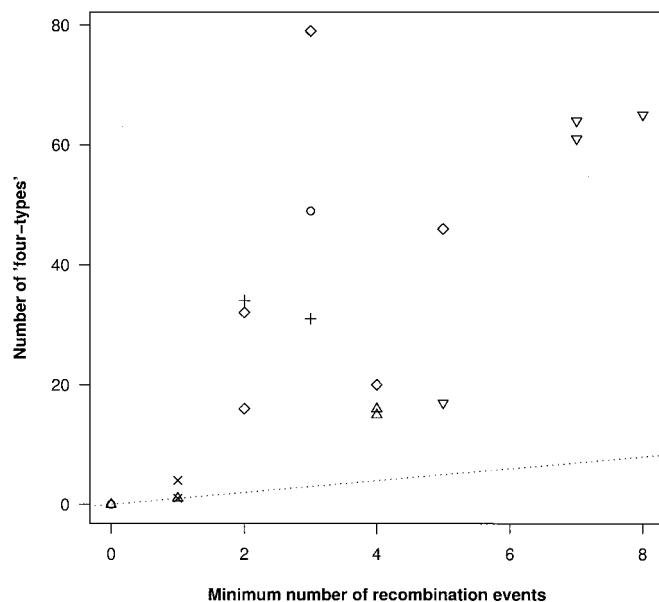


**Fig. 3.** Estimates of Tajima's  $D$  against pairwise diversity for viruses sequenced from each pulp. The dashed line indicates significantly ( $P$ , approximately less than 0.05) negative values of  $D$  ( $< -1.84$ ), whereas the dotted line indicates  $D = 0$  (the expected value under a constant population evolving neutrally). Data from different individuals are represented by different symbols (see Fig. 2 legend). There is a significant ( $P < 0.001$ ) correlation between Tajima's  $D$  and pairwise diversity.

of Tajima's  $D$ , corrected for heterogeneity in the substitution rate along sites (47), were more strongly negative (not shown). No significant bottleneck effect was found when within-pulp diversity is high, consistent with the metapopulation model, where high levels of diversity are generated by multiple, genetically divergent founders seeding pockets of uninfected cells.

Metapopulation dynamics may also play a role in setting the effective recombination rate, as recombinant viruses are produced only when multiple viruses infect the same cell (which occurs at a local spatial scale) and can be detected only when these viruses are genetically different, which under our metapopulation model is affected by the processes of birth and death of subpopulations. We first tested for the presence of recombinant viruses within each individual as a whole by looking for long regions in which pairs of sequences were unusually similar ("inner" fragments), suggesting recombination between them or their ancestors, or unusually different ("outer" fragments), suggesting recombination with a sequence that is either not represented in the sample, or where further mutation or recombination has obscured the relationship. Fragments could be identified in four of six patients at the  $P = 0.05$  level (B, L, M, and P), and in one additional patient (S) at the  $P = 0.1$  level. We hypothesized that within each individual, more recombination events would be observed in white pulps with higher diversity, as multiple genetically divergent founders increase the opportunities for new genetic variants to be produced locally by recombination. To test this hypothesis, we first estimated the minimum number of recombination events within each pulp from the number of pairs of variable sites where all four combinations of mutations were observed ("four-types"). Four-types were identified in 17/20 pulps. Although these combinations can also be produced by independent substitutions at each site, rather than by recombination between the sites, the number of four-types was usually much higher than the minimum number of recombination events (median 8.1-fold higher), with 13/17 pulps where four-types were at least 4-fold higher (Fig. 4). The large differ-





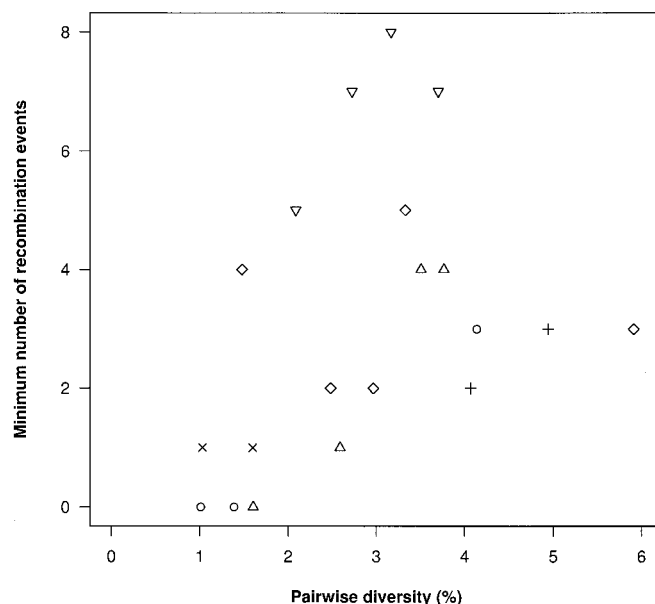
**Fig. 4.** Estimates of the number of “four-types,” i.e., the number of pairs of sites where all four combinations of two mutants are present, against the minimum number of recombination events required to produce them, plotted by pulp for each patient (symbols as in Fig. 2 legend). The number of recombination events that need to be invoked to explain the data is much lower than the number of four-types (as shown by the large number of points above the dotted line, which has an intercept 0 and slope 1).

ence in the number of four-types and the minimum number of recombination events in the majority of pulps analyzed is a consequence of the clustered distribution of the “four-types” along the sequence, which is more likely to arise under recombination than under convergent evolution. There was a significant positive correlation between pairwise diversity and the minimum number of recombination events estimated from the pattern of four-types (Spearman’s rank correlation = 0.48,  $P < 0.05$ ; Fig. 5), consistent with the hypothesis that the effective recombination rate may be affected by the number of founding viruses and subpopulation turnover via levels of within pulp diversity.

## Discussion

The assumption that HIV populations are panmictic is inconsistent with the pattern of genetic variation in solid tissues. Rather, they support a “metapopulation” model, as originally defined in terms of population turnover as a “population of populations,” with an age structure established by the birth of new populations through colonization and their death through extinction (39). Large differences between individuals in the level of viral genetic structuring within the spleen can arise because of relatively small differences in the number of founders, with high levels of population structure, low levels of diversity within subpopulations, and a lesser role of recombination arising when the number of founders is low. The metapopulation concept, in a more general sense where a population is made up of discrete subunits that exchange genes, has been invoked in a large number of studies in ecology, conservation biology, and population genetics (53). The impact of metapopulation processes on genetic diversity has been discussed recently in an excellent review by Pannell and Charlesworth (54).

A comparison of the amount of genetic differentiation between viruses from different white pulps of the spleen with the total amount of viral genetic variation clearly demonstrated spatial structure, although the extent of population structuring



**Fig. 5.** Estimates of minimum number of recombination events against pairwise diversity for viruses sequenced from each pulp. Data from different individuals are represented by different symbols (see Fig. 2 legend). There is a significant ( $P < 0.05$ ) positive correlation between the minimum number of recombination events and pairwise diversity.

varied between individuals. Although the genetic diversity, both within subpopulations and as a whole, is likely to be lower than their equilibrium values [because of the relatively genetically homogeneous viral population at the initial bottleneck of infection (6, 55–58)], the ratio of the two, as measured by  $F_{ST}$ , reaches its equilibrium value relatively rapidly, and hence estimates of  $F_{ST}$ , and our conclusions based on them are not expected to be greatly affected by the bottleneck of infection.

The use of Tajima’s  $D$  statistic as a summary statistic describing the shape of the viral phylogeny within pulps showed that low genetic diversity was associated with low values of  $D$ , which we concluded were because of founder effects. Although selection can also result in deviations of  $D$  from 0, selection within pulps is likely to be weak because of the very small within-pulp population size, and hence any significant departures are likely to reflect founder effects. Interestingly, when Tajima’s  $D$  was calculated for the pooled sequences from each individual, low levels of genetic differentiation were associated with significantly negative values of  $D$ , especially when rate heterogeneity in mutation rate along the sequence was taken into account (Table 2). There are several plausible explanations. First, the age structure of subpopulations, which arises as a consequence of founding and turnover, has been hypothesized to result in negative values of  $D$  when there is sufficient within-population diversity (59), although the size of this effect requires knowledge of the distribution of coalescence times under a metapopulation model. This is currently an open problem, although progress is being made in this area.<sup>||</sup> Second, the viruses in different pulps in the spleen may themselves have gone through a bottleneck. In virus isolated from peripheral blood, it has been shown that a bottleneck effect is present early in infection that wanes as infection progresses (60), and this may reflect a similar process in solid tissue. Third, negative values of  $D$  can arise when neutral variation is linked to selectively advantageous mutations that

<sup>||</sup>Wakeley, J., 8th Annual Discussion Meeting on HIV Dynamics and Evolution, Paris, April 2001.

**Table 2. Estimates of average pairwise diversity, number of mutations, and Tajima's *D* (no rate heterogeneity) and Misawa and Tajima's *D* (corrected for rate heterogeneity) for the pooled sequences**

Patient	No. of sequences (length in base pairs)	Average no. of differences (%)	Number of mutations	Tajima's <i>D</i>	Misawa and Tajima's <i>D</i> ( $\alpha = 0.3$ )
B	55 (257 bp)	10.63 (4.14)	48	0.045 ( $P = 0.53$ )	0.154 ( $P = 0.57$ )
L	81 (217 bp)	8.68 (4.00)	75	-1.414 ( $P = 0.072$ )	-1.678 ( $P = 0.036$ )
M	38 (268 bp)	11.98 (4.47)	79	-1.328 ( $P = 0.091$ )	-1.547 ( $P = 0.054$ )
N	30 (265 bp)	3.524 (1.33)	33	-2.099 ( $P = 0.007$ )	-2.237 ( $P = 0.003$ )
P	90 (285 bp)	11.076 (3.89)	110	-1.633 ( $P = 0.04$ )	-1.932 ( $P = 0.015$ )
S	77 (229 bp)	6.734 (2.94)	82	-2.001 ( $P = 0.011$ )	-2.289 ( $P = 0.003$ )

sweep through the population (61). Fourth, negative values of *D* may arise because of purifying selection on deleterious mutations (26). Although selection on a single deleterious mutation has little or no effect on phylogenetic tree shape (62), this effect may be greater if selection acts on multiple mutations.

Metapopulation dynamics are likely to affect the impact of recombination on viral genetic variation through its effects on subpopulation diversity. We demonstrated a positive correlation across pulps between pairwise diversity and the estimated minimum number of recombination events based on the pattern of "four-types," pairs of sites where all four combinations of two mutants are present. We interpreted this correlation as showing that low diversity arising because of founder effects restricts the impact of recombination, which occurs on a local rather than a global scale. Although "four-types" can also be generated by convergent evolution, their clustered spatial pattern along the sequence suggests that they were more likely to be generated by a few recombination events rather than by a large number of multiple substitutions. A more rigorous model-based test of whether these patterns arise because of recombination rather than multiple substitutions requires knowledge of the distribution of coalescent times under a metapopulation model. Although a positive correlation between genetic diversity and the number of recombination events could be generated if recombinants formed part of a founding viral population rather than being produced locally, this would also require the presence of the parental strains in the founding population. This seems biologically unlikely, and the number of four-types is likely to reflect recombination events, which occur after founding of the subpopulation.

Although spatial models have been applied to the question of HIV population structure, these models have been of the "island" type, which does not allow for extinction of subpopulations, and where dissemination of virus is assumed to occur only between extant subpopulations, such as between different parts of the brain (63) and between the plasma and cerebrospinal fluid.\*\* Although this may be a reasonable assumption when looking at large-scale population structure (as large subpopulations are unlikely to go extinct), or when looking at infection of glial cells in the brain (which undergo slow turnover), turnover of productively infected T cells in solid lymphoid tissue is high because of local exhaustion of target cells and/or to the infiltration of HIV-specific cytotoxic T cells (37, 38) and with low levels of mixing will lead to extinction of foci of infection. Under the island model, the large differences in  $F_{ST}$  seen between individuals would have to be generated by large differences in the migration rate, rather than subtle differences in the number of founders, as in the metapopulation model. The low level of diversity and the presence of founder effects and local recombinant viruses in subpopulations of infected cells suggest they

have been recently generated, rather than being set up early in infection, with continuous migration between subpopulations.

Small-scale spatial variation of HIV has important implications for within-host HIV evolution. The effective population size,  $N_e$ , under a metapopulation model is lower than the actual number of productively infected cells,  $N_T$ , whereas  $N_e$  is greater than  $N_T$  under the island model (64). For the metapopulation model, the ratio of effective population size to total population size is given by

$$\frac{N_e}{N_T} = \frac{k(1-x)(1+N_{Sx})}{N_{Sx}(3x+2N_{Sx}-1)} \approx \frac{k}{N_S} \frac{1}{2V}. \quad [2]$$

The approximation is for moderate amounts of turnover,  $x \gg 1/N_S$ , and  $V$  is the variance in fitness between subpopulations, which for our simple metapopulation model is given by  $V = x/(1-x)$ . Eq. 2 clearly illustrates that the more dramatic the founder effect and the higher the variance in fitness between subpopulations, the lower the effective population size is relative to the census size. Thus, if  $k = 2$ ,  $x = 0.1$ , and  $N_S = 100$ ,  $N_e$  is approximately an order of magnitude lower than the actual population size and would be even lower if there were differences in productivity between subpopulations, which would increase  $V$  further.

Metapopulation structure could play a central role in the evolution of HIV populations within infected patients via its effect on  $N_e$ . Population subdivision can lead to a dramatic reduction in  $N_e$  when subpopulations turn over, because of the disproportionately high fitness of founders (64). Hence, local extinction of subpopulations could contribute to the highly variable way in which drug resistance evolves in different individuals (25, 26), even though the absolute number of infected cells is high ( $10^7$ – $10^8$ ) (4). Rouzine and Coffin (28) estimated the effective population size of HIV at  $10^4$ – $10^5$  based on the frequency of all four combinations of mutations at variable sites in virus isolated from peripheral blood mononuclear cells, on the assumption that these combinations are produced by independent substitutions rather than recombination. We have demonstrated recombination at the local level, and although it is not known how recombination at a subpopulation level translates to the effective recombination rate in the body as a whole, estimates of  $N_e$  based on the level of linkage disequilibrium are likely to be overestimates. Nevertheless, metapopulation structure is unlikely to be the sole cause of differences between the actual and effective population sizes as estimates of  $N_e$  based on the level of diversity (24, 36) and the rate of divergence of the C2-V3 region of the *env* gene (27) are of the order  $10^3$ – $10^4$ , which is lower than could realistically be produced under our simple metapopulation model. Other factors, such as the process of "genetic draft" (61), where a reduction in effective population size arises because of selection on linked viral mutations, may play an important role.

The consequences of a metapopulation of infected cells extend beyond the impact on effective population size. Although they

\*\*Wang, Y., Marra, C. M., Learn, G. H., Rodrigo, A. G., Collier, A. C., Coombs, R. W., He, X., Zhao, H., & Mullins, J. I., Abstract no. 307, 7th Conference of Retroviruses and Opportunistic Infections, San Francisco, January 2000.

did not use the term “metapopulation,” Grossman *et al.* (65, 66) suggested that local short-lived bursts of replication could aid persistence of the virus during highly active antiretroviral therapy (HAART), and that changes in the dynamics of these bursts during therapy could account for the “biphasic” decay of virus in the blood observed during HAART. Metapopulation structure is imposed by the spatial structure of lymphoid tissues, which has been hypothesized to play a key role in determining how the immune system “chooses” an appropriate response against parasites (67), and the compartmentalization of specific immune cells in HIV infection (37) is also likely to have important implications for the way in which viral immune escape

mutants evolve. Like viral infection of target cells, many other processes that play a role in viral replication, such as the effects of cytokines and bystander effects, (68, 69) are also local in nature, and an understanding of their role in viral dynamics may benefit from consideration of virally infected cells as a “population of populations.”

Thanks go to John Wakeley and two anonymous referees for comments on the manuscript. This work was funded in part by a Medical Research Council Fellowship (681/298) to S.D.W.F. and by an unrestricted gift from Roche Molecular Systems, Alameda, CA.

- Mansky, L. M. & Temin, H. M. (1995) *J. Virol.* **69**, 5087–5094.
- Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M. & Markowitz, M. (1995) *Nature (London)* **373**, 123–126.
- Wei, X., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., Lifson, J. D., Bonhoeffer, S., Nowak, M. A., Hahn, B. H., *et al.* (1995) *Nature (London)* **373**, 117–122.
- Chun, T. W., Carruth, L., Finzi, D., Shen, X., DiGiuseppe, J. A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T. C., *et al.* (1997) *Nature (London)* **387**, 183–188.
- Simmonds, P., Zhang, L. Q., McOmish, F., Balfe, P., Ludlam, C. A. & Leigh Brown, A. J. (1991) *J. Virol.* **65**, 6266–6276.
- Wolfs, T. F. W., Zwart, G., Bakker, M. & Goudsmit, J. (1992) *Virology* **189**, 103–110.
- Cichutek, K., Merget, H., Norley, S., Linde, R., Kreuz, W., Gahr, M. & Kurth, R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7365–7369.
- Bonhoeffer, S., Holmes, E. C. & Nowak, M. A. (1995) *Nature (London)* **376**, 125.
- Leigh Brown, A. J. & Cleland, A. (1996) *AIDS* **10**, 1067–1073.
- Liu, S. L., Schacker, T., Musey, L., Shriner, D., McElrath, M. J., Corey, L. & Mullins, J. I. (1997) *J. Virol.* **71**, 4284–4295.
- McDonald, R. A., Mayers, D. L., Chung, R. C. Y., Wagner, K. F., RattoKim, S., Bix, D. L. & Michael, N. L. (1997) *J. Virol.* **71**, 1871–1879.
- Yamaguchi, Y. & Gojibori, T. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 1264–1269.
- Nielsen, R. & Yang, Z. (1998) *Genetics* **148**, 929–936.
- Yang, Z. H., Nielsen, R., Goldman, N. & Pedersen, A. M. K. (2000) *Genetics* **155**, 431–449.
- Yamaguchi-Kabata, Y. & Gojibori, T. (2000) *J. Virol.* **74**, 4335–4350.
- Frost, S. D. W., Günthard, H. F., Wong, J. K., Havlir, D., Richman, D. D. & Leigh Brown, A. J. (2001) *Virology*, in press.
- Boucher, C. A. B., O’Sullivan, E., Mulder, J. W., Ramautarsing, C., Kellam, P., Darby, G., Lange, J. M. A., Goudsmit, J. & Larder, B. A. (1992) *J. Infect. Dis.* **165**, 105–110.
- Condra, J. H., Holder, D. J., Schleif, W. A., Blahy, O. M., Danovich, R. M., Gabryelski, L. J., Graham, D. J., Laird, D., Quintero, J. C., Rhodes, A., *et al.* (1996) *J. Virol.* **70**, 8270–8276.
- Molla, A., Korneyeva, M., Gao, Q., Vasavanonda, S., Schipper, P. J., Mo, H. M., Markowitz, M., Chernyavskiy, T., Niu, P., Lyons N., *et al.* (1996) *Nat. Med.* **2**, 760–766.
- Crandall, K. A., Kelsey, C. R., Imamichi, H., Lane, H. C. & Slazman, N. P. (1999) *Mol. Biol. Evol.* **16**, 372–382.
- Havlir, D. V., Eastman, S., Gamst, A. & Richman, D. D. (1996) *J. Virol.* **70**, 7894–7899.
- Eastman, P. S., Mittler, J., Kelso, R., Gee, C., Boyer, E., Kolberg, J., Urdea, M., Leonard, J. M., Norbeck, D. W., Mo, H. & Markowitz, M. (1998) *J. Virol.* **72**, 5154–5164.
- Frost, S. D. W., Nijhuis, M., Schuurman, R., Boucher, C. A. B. & Leigh Brown, A. J. (2000) *J. Virol.* **74**, 6262–6268.
- Nijhuis, M., Boucher, C. A. B., Schipper, P., Leitner, T., Schuurman, R. & Albert, J. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 14441–14446.
- Leigh Brown, A. J. & Richman, D. D. (1997) *Nat. Med.* **3**, 268–271.
- Leigh Brown, A. J. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 1862–1865.
- Rodrigo, A. G., Shpaer, E. G., Delwart, E. L., Iversen, A. K. N., Gallo, M. V., Brojatsch, J., Hirsch, M. S., Walker, B. D. & Mullins, J. I. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2187–2191.
- Rouzine, I. M. & Coffin, J. M. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 10758–10763.
- Ribeiro, R. M. & Bonhoeffer, S. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 7681–7686.
- Coffin, J. M. (1995) *Science* **267**, 483–489.
- Epstein, L. G., Kuiken, C., Blumberg, B. M., Hartman, S., Sharer, L. R., Clement, M. & Goudsmit, J. (1991) *Virology* **180**, 583–590.
- Delassus, S., Meyerhans, A., Cheynier, R. & Wain-Hobson, S. (1992) *Virology* **188**, 811–818.
- Hughes, E. S., Bell, J. E. & Simmonds, P. (1997) *J. Gen. Virol.* **78**, 2871–2882.
- Wong, J. K., Ignacio, C. C., Torriani, F., Havlir, D., Fitch, N. J. S. & Richman, D. D. (1997) *J. Virol.* **71**, 2059–2071.
- Schrager, L. K. & D’Souza, M. P. (1998) *J. Am. Med. Assoc.* **280**, 67–71.
- Kashuba, A. D. M., Dyer, J. R., Kramer, L. M., Raasch, R. H., Eron, J. J. & Cohen, M. S. (1999) *Antimicrob. Agents Chemother.* **43**, 1817–1826.
- Cheynier, R., Henrichwark, S., Hadida, F., Pelletier, E., Oksenhendler, E., Autran, B. & Wain-Hobson, S. (1994) *Cell* **78**, 373–387.
- Gratton, S., Cheynier, R., Dumaourier, M.-J., Oksenhendler, E. & Wain-Hobson, S. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 14566–14571.
- Levins, R. (1969) *Bull. Ent. Soc. Am.* **5**, 237–240.
- Thomson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
- Wright, S. (1951) *Ann. Eugen.* **15**, 323–354.
- Schneider, S., Roessli, D. & Excoffier, L. (2000) *ARLEQUIN: A Software for Population Genetics Analysis* (Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva).
- Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York) pp. 21–132.
- Rozas, J. & Rozas, R. (1999) *Bioinformatics* **15**, 174–175.
- Tajima, F. (1989) *Genetics* **123**, 585–595.
- Simonsen, K. L., Churchill, G. A. & Aquadro, C. F. (1995) *Genetics* **141**, 413–429.
- Misawa, K. & Tajima, F. (1997) *Genetics* **147**, 1959–1964.
- Leitner, T., Kumar, S. & Albert, J. (1997) *J. Virol.* **71**, 4761–4770.
- Sawyer, S. A. (1999) *GENECONV: A Computer Package for the Statistical Detection of Gene Conversion* (Department of Mathematics, Washington University, St. Louis, MO).
- Hudson, R. R. & Kaplan, N. L. (1985) *Genetics* **111**, 147–164.
- Slatkin, M. (1977) *Theor. Popul. Biol.* **12**, 253–262.
- Pannell, J. R. & Charlesworth, B. (1999) *Evolution (Lawrence, KS)* **53**, 664–676.
- Hanski, I. & Gilpin, M. E., eds. (1997) *Metapopulation Biology: Ecology, Genetics, and Evolution* (Academic, San Diego).
- Pannell, J. R. & Charlesworth, B. (2000) *Philos. Trans. R. Soc. London B* **355**, 1851–1864.
- McNearney, T., Hornickova, Z., Markham, R., Birdwell, A., Arens, M., Saah, A. & Ratner, L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10247–10251.
- Zhu, T. F., Hunag, Y. X., Wang, N., Connor, R., Cao, Y. Z. & Ho, D. D. (1993) *J. Cell Biochem.* **S17E**, 25.
- Zhang, L. Q., Mackenzie, P., Cleland, A., Holmes, E. C., Leigh Brown, A. J. & Simmonds, P. (1993) *J. Virol.* **67**, 3345–3356.
- Long, E. M., Martin, H. L., Kreiss, J. K., Rainwater, S. M. J., Lavreys, L., Jackson, D. J., Rakwar, J., Mandaliya, K. & Overbaugh, J. (2000) *Nat. Med.* **6**, 71–75.
- Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993) *Genetics* **134**, 1289–1303.
- Crandall, K. A., Vasco, D. A., Posada, D. & Imamichi, H. (1999) *AIDS* **13**, S39–S47.
- Gillespie, J. H. (2000) *Genetics* **155**, 909–919.
- Golding, G. B. (1997) in *Progress in Population Genetics and Human Evolution*, eds. Donnelly, P. & Tavaré, S. (Springer, New York), pp. 271–285.
- Shapshak, P., Segal, D. M., Crandall, K. A., Fujimura, R. K., Zhang, B. T., Xin, K. Q., Okuda, K., Petit, C. K., Eisdorfer, C. & Goodkin, K. (1999) *AIDS Res. Hum. Retr.* **15**, 811–820.
- Whitlock, M. C. & Barton, N. H. (1997) *Genetics* **146**, 427–441.
- Grossman, Z., Feinberg, M. B. & Paul, W. E. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 6314–6319.
- Grossman, Z., Polis, M., Feinberg, M. B., Grossman, Z., Levi, I., Jankelevich, S., Yarchoan, R., Boon, J., de Wolf, F., Lange, J. M., *et al.* (1999) *Nat. Med.* **5**, 1099–1104.
- Segel, L. A. & Bar-Or, R. L. (1999) *J. Immunol.* **163**, 1342–1349.
- Frost, S. D. W. & Michie, C. A. (1996) *Trends Microbiol.* **4**, 77–82.
- Krakauer, D. C. & Nowak, M. A. (1999) *Proc. R. Soc. London Ser. B* **266**, 1069–1075.